The Royal Society of Medicine

Journal of the Royal Society of Medicine; 2015, Vol. 108(8) 325–330
DOI: 10.1177/0141076815596690

Why did I become a clinician-trialist?

David Sackett†
Former Foundation Professor of Clinical Epidemiology, McMaster University, Hamilton, Ontario, Canada; and Foundation Director, Centre for Evidence-Based Medicine, Oxford University, UK
Corresponding author: Iain Chalmers. Email: ichalmers@jameslindlibrary.org

The seeds of iconoclasm

In 1956, I went to the University of Illinois, College of Medicine to be taught how to become a physician. Despite the College's well-deserved reputation, a recent therapeutic scandal was smouldering – and occasionally bursting into flames. The university's Vice-President-Director was Dr Andrew Ivy, a famous gastrointestinal physiologist. He had represented the American Medical Association at the medical trials of Nazis in Nuremberg, and he subsequently became Executive Director of the National Advisory Cancer Council and a director of the American Cancer Society. Not long before my arrival at the College of Medicine, Dr Ivy had been accused of fraudulently claiming efficacy for a quack cancer remedy, Krebiozen, which turned out to be nothing more than creatine.[1] None of my teachers (some of whom were involved in attempts to resolve the dispute) ever spoke about the scandal, but it had generated an atmosphere of skepticism towards authority figures, and this had fostered iconoclasm around the place, which appealed to me.

By 1959, I had become a final-year medical student, and I once found myself responsible for a teenager who had been admitted to a medical ward with hepatitis. After a few days of enforced total bed rest – the standard management of the condition – his spirits and energy returned and he asked me to let him get up and around. I felt I needed to have a look at relevant evidence to guide my response to his request. I went to the library and came across a remarkable report[2] for which the lead author was Tom Chalmers.[3] A meticulously conducted randomised trial had made clear that there was no good evidence to justify requiring hepatitis patients to remain in bed after they feel well. Armed with this evidence, I convinced my supervisors to let me apologise to my patient and encourage him to be up and about as much as he wished. His subsequent clinical course was uneventful. That report of a (factorial) randomised trial challenging the validity of two standard treatments for hepatitis – bed rest and low fat diet – helped to change my career.[4]

During my postgraduate training in internal medicine, the better I became at diagnosing my patients' illnesses, the more frustrated I became at my profession's collective ignorance about how I should treat them, or whether I should treat them at all. I came to the conclusion that there were four things wrong with the way that the experts were using their clinical observations to decide whether a treatment did more good than harm. More precisely, I was worried that these four 'wrongs' destroyed our ability to make 'fair comparisons' of the effects of different treatments. The validation of these worries both initiated and reinforced my decision to devote most of my career to randomised trials.

Worry #1: I became worried that clinicians might preferentially give new treatments to patients with better prognoses

One of my 'rotations' as a first-year medical resident was the Admitting Clinic. I evaluated referrals from all over Illinois to assess whether they would be 'good teaching cases' for the medical and surgical services at our Research and Educational Hospital. My surgical resident colleague explained to me that they had two 'general surgery' services, and that they evaluated innovative operations by performing them on the 'A Service' (where he scrubbed) while continuing to perform standard operations on the 'B Service'.

Although a perfect setting for randomisation, when we examined a patient and found them suitable for one of their comparative studies, my surgical colleague decided where they went. Over time, I became convinced that he was preferentially admitting eligible surgical patients with sounder hearts, healthier lungs, and higher haematocrits to receive the new, promising operations on his 'A Service.' Thus, sensitised, I began to pay more attention to the therapeutic recommendations for new, untested treatments I received from my senior consultants. Again, I concluded that, within the same illness, it

†The author died recently

was my healthier patients whom they considered 'good candidates' for the latest, untested treatment.

It was decades later that I was introduced to a very telling confirmation of this first concern. In New York City in the 1930s, babies born into households that included members with pulmonary tuberculosis were at high risk of dying from the disease before their first birthday. Although the BCG vaccine was already in use and touted as protecting such infants, a New York City public health team that included Margaret Sackett was skeptical about these claims. I do not know whether we are related, but I hereby claim to be her long-lost nephew because she carried out two BCG 'trials'.[5] In the first 'trial', public health physicians were assigned batches of at-risk newborns and told: 'vaccinate half of them'. The results were spectacular: the risk of dying before their first birthday was reduced by 80% among vaccinated babies.

In the second 'trial', however, the decision about whom to vaccinate was taken out of the physicians' hands and was determined by 'drawing lots', generating a fair assessment of BCG efficacy. The results were no less spectacular, but in this case quite 'negative': the risk of dying before their first birthday was identical between vaccinated and non-vaccinated babies. This presented the opportunity to determine how the physicians in the first trial (told to 'vaccinate half of them') made the decision to vaccinate some babies but not others. This inquiry revealed that they were more likely to vaccinate babies who were headed for wealthier, less crowded households whose family members had less severe tuberculosis. The BCG-inoculated babies had better prognoses before they were vaccinated!

Clinicians often do preferentially treat patients with better prognoses. That's why our randomised controlled trials employed the 'fair comparison' strategies of random allocation and concealment (from treating clinicians) of the treatment that was destined to be given to the patient they were considering enrolling in randomised controlled trials.[6]

Worry #2. I became worried that patients compliant with treatment instructions might have better prognoses, regardless of their treatment

My first five clinical years as student and postgraduate trainee gave me the opportunity to observe and contribute to the care of a few hundred patients. I kept an irregular list of their treatments, clinical courses and outcomes folded into my copy of *Harrison's Textbook of Medicine*. As these notes accumulated, two perplexing conclusions emerged.

First, I was surprised to discover that only about half of my patients refilled their prescriptions regularly and took their medicine (it was already 'common knowledge' that we physicians were poor compliers, but we'd naively thought our patients were much better). Some patients simply disappeared, and those that returned to our clinic continued their poor compliance despite our exhortations, and often succumbed to their illnesses.

Second, those of my patients who refilled their prescriptions on time and appeared compliant not only had better prognoses but appeared to improve regardless of whether, on the one hand, my treatments were supported by strong evidence (for example, the early trials in complicated severe hypertension), or by little or no evidence (for example, the contemporary treatments for coronary heart disease) on the other. Looking more closely, I noted that these 'compliant' patients were also less likely to be smokers, heavy drinkers, or overweight. Finally, and harking back to my first 'worry', they were often the patients who my seniors picked as 'good candidates' for new, untested treatments. On the basis of the foregoing, I began to worry whether high compliance might be a 'marker' for rosier prognoses, regardless of therapy.

Confirmation of this 'worry' had to wait for compelling examples of this phenomenon in analyses of placebo groups in randomised trials. For example, when I was a house officer in Buffalo in 1966, I had entered patients who had had a heart attack into a trial comparing one of several of that decade's lipid-lowering agents with placebo. The Coronary Drug Project Research Group[7] was hard-pressed to find a drug that made any difference. For example, the five-year mortality for participants randomised to clofibrate (20%) was no better than for those randomised to placebo (21%).

The hopes of the trialists rose when they noted that one-third of clofibrate-assigned patients were taking less than 80% of their assigned treatment, and they decided that a better measure of clofibrate's efficacy would be to compare the mortality of clofibrate non-compliers with that of the majority who were taking 80% or more of the prescribed drug. The results were (temporarily) encouraging: good 'adherers' to clofibrate had substantially lower five-year mortality than did poor adherers to clofibrate (0.15 vs. 0.246; relative risk ratio = 39%; $z = -3.86$; $P = 0.00011$).

However, the hero-statistician of the trial, Paul Canner, carried out a similar analysis for participants who did and did not take their *placebos* as instructed. He showed an even stronger association between compliance effect and mortality (0.151 vs.

0.282; relative risk ratio = 46%; $z = -8.12$; $P = 0.00000000000000047$), implying that one premature death would be prevented for every 10 patients who took their placebo faithfully!

In a major contribution to our (?non-) understanding of the 'compliance-effect', the research team showed that the increased risk of death among poor placebo compliers could not be accounted for by taking account of 40 baseline characteristics associated with five-year mortality, the characteristics that one might insert these days into a 'propensity score' in an attempt to create comparable groups using statistical adjustments.[8] After this 'propensity score correction', the relative risk reduction of 46% only fell to 36%, the z-score from -8.12 to -5.78 and the P value from 0.00000000000000047 to a still-overwhelming 0.00000000073.

The investigators concluded:

> These findings and various other analyses of mortality in the clofibrate and placebo groups of the project show the serious difficulty, if not impossibility, of evaluating treatment efficacy in subgroups determined by patient responses (e.g., adherence [to treatment], or cholesterol change) to the treatment protocol after randomization.

Compliant patients do have better prognoses, regardless of their prescribed treatment (as long as it is not inherently toxic). Thus, (inappropriately called) 'per-protocol' analyses confined to compliant patients are inherently invalid. That is why our randomised controlled trials have employed the 'fair comparison' strategies of unobtrusive compliance measures, intention-to-treat analyses, and keeping track of everybody who enters them. Walsh et al.[9] have documented that over 50% of 'positive' randomised controlled trials in leading journals have losses to follow-up that exceed the fragility of their positive result. I recently toted up the losses to follow-up among the >12,000 participants in the trials in which I have been a principal investigator and was cheered to find that it was only 0.4%.

Worry #3. I became worried that patients who liked their treatment might report spuriously better outcomes

As clinical clerks on the internal medicine service, we were encouraged to read every week's issues of the *Journal of the American Medical Association (JAMA)* and the *New England Journal of Medicine (NEJM)*. For example, in May of 1959, we learned from JAMA about the first few successful cardiopulmonary resuscitations, and how the active ingredient in the Sabin polio vaccine rapidly spreads throughout an institutional population. The *NEJM* told us how to select patients for 'definitive' surgery for their duodenal ulcers, and how we could obtain rapid polio immunisation by injecting 10 mL of the Salk vaccine.

But the paper in the *NEJM* that made the greatest, lasting impression on me was a report from a surgeon. Cobb et al.[10] had randomised and blinded patients who were so seriously limited by angina that the majority were unemployed. Randomised to what? In the decade before their trial, thousands of angina patients had undergone the 'miracle operation' of internal mammary artery ligation (based on the theory that blood previously coursing down these arteries would be partially redistributed to the coronary circulation). As reported in *Readers' Digest* in July 1957[11]: 'complete or partial relief from the pain that accompanies the major types of heart disease has been obtained in nearly 80% of the several hundred operations performed to date'. This simple operation (done under local anesthesia in just a few minutes) became so popular that one wag suggested that: 'It is, perhaps, surprising that between 1955 and 1960 there were still patients with angina whose mammary arteries were not ligated'. Indeed, all three of the patients I had examined who had surgical scars over their ribs claimed their operations had improved or relieved their angina. Thus, although in Cobb's randomised trial, 'subjects were informed of the fact that this procedure had not been proved to be of value, ... many were aware of the enthusiastic report published in the Readers' Digest'.

In Cobb's trial, a screen prevented patients from seeing what was happening as their internal mammary arteries were surgically exposed. After a ligature had been placed loosely around these arteries, the surgeon was handed a 'randomly selected envelope' which contained a card instructing him either to tie off the arteries or to remove the loose ligature and leave the arteries alone. Thus, the patients had neither the choice nor the knowledge of whether their arteries were ligated.

During their 3- to 15-month follow-up by physicians who were kept unaware of the group to which each trial participant had been assigned (ligation or not), some spectacular results were documented: for example, Case #4, who had previously been unable to work because of his angina, reported almost instant relief and was able to return to work (in fact, his arteries had not been ligated).

On the other hand, 'The average improvement was 32% for the ligated patients and 43% for those whose internal mammary arteries were not ligated'. The trialists concluded: 'Bilateral skin incisions in the second intercostal space seem to be at least as

effective as internal-mammary-artery ligation in the therapy of angina pectoris'.

Although internal mammary ligation rapidly disappeared after this and a second randomised trial was reported,[12,13] this 'positive expectation bias' has continued to haunt attempts to critically appraise therapeutic fads to the present day, as we continue to debate the efficacy of 'liberation therapy' for patients with multiple sclerosis

Patients who like their treatment do report better outcomes unrelated to the true efficacy of their treatments. That's why our randomised controlled trials employed (whenever possible, and it is possible more than detractors might think) blinding of trial patients to their treatments, 'hard' outcomes such as total mortality, and the 'blind' adjudication of softer outcomes.

Worry #4. I was worried that clinicians who liked their treatment might report spuriously better outcomes

The internal mammary ligation fiasco also hardened my worry that physicians writing prescriptions might be as guilty of over-reporting their favourable effects as the patients who filled and consumed them. Although the James Lind Library notes that the need for the blind assessment of treatment effects was emphasised many years before I was born,[14] the hardest evidence that clinicians who like their treatments report spuriously better outcomes comes from far more recent randomised controlled trials.

For example, in a promising placebo-controlled Canadian randomised controlled trials of weekly plasma exchange, prednisone and cyclophosphamide among patients with multiple sclerosis, two sets of neurologists were asked to determine treatment responses at 6, 12 and 24 months.[15] Neurologists who were blind to the treatments reported no difference in outcomes among the treatment groups at any time. However, unblinded neurologists reported statistically significantly improved outcomes for patients receiving triple therapy at all three follow-up assessments.

Clinicians who like their treatment do report spuriously better outcomes. That is why our randomised trials use blinded outcome assessors whenever we can, draw conclusions from 'hard' outcomes if possible, and blinded adjudication of softer outcomes.

Concluding reflections

Randomised trials are not always possible for investigating putative effects of treatment: but numerous actual examples show that they are more often an option (such as the trial by Cobb and his colleagues described above) than many people believe. The main precondition seems often to be the professional humility to admit that, on the basis of the evidence available, we are uncertain whether a treatment is more likely to do good than harm, and the need to use reliable research to identify its effects.

On the other hand, for investigating the harmful effects of treatment of some possible treatment effects – particularly alleged rare adverse effects – observational data from case-control and cohort analytic research will be required.[16–19] The 1980s saw active debates about the validity of observational studies for investigating the possible adverse effects of drugs, and I contributed to a meeting chaired by Michel Ibrahim (Appendix), which discussed and debated the conflicting views about the validity of this study design. My contribution was to compile a catalogue of the biases that might need to be taken into account in evaluating observational data.[20] One of the effects of my contribution was to misinform Big Pharma that I could be a hired gun to trash observational studies revealing the lethality of their drugs!

The biases that I identified in *Bias in Analytic Research* have not disappeared with the passage of time. As I witness the emerging era of Comparative Effectiveness Research, I have not encountered convincing examples in which the proponents of observational studies of efficacy (as distinct from adverse effects) have developed strategies and tactics for avoiding and/or overcoming the four worries that forced me into hard randomised trial labour for the past 48 years. Indeed, I'm curious about how they will (and could) tell whether they've avoided or solved them.

Declarations

Competing interests: None declared

Funding: None declared

Ethical approval: Not applicable

Guarantor: DS

Contributorship: Sole author

Acknowledgement: I am grateful to Michel Ibrahim for reminding me of some of the features of the conference on case-control studies that he organised in 1978 (see Appendix).

Provenance: Invited contribution from the James Lind Library

References

1. Ivy AC. Krebiozen. *Science* 1951; 114: 285–286.
2. Chalmers TC, Eckhardt RD, Reynolds WE, Cigarroa JG, Deane N, Reifenstein RW, et al. The treatment of acute infectious hepatitis. Controlled studies of the effects of diet, rest, and physical reconditioning on the acute course of the disease and on the incidence of relapses and residual abnormalities. *J Clin Investig* 1955; 34: 1163–1235.

3. Dickersin K and Chalmers F. Thomas C Chalmers (1917–1995): a pioneer of randomized clinical trials and systematic reviews. *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2014, accessed 2 July 2015).
4. Sackett DL. A 1955 clinical trial report that changed my career. *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2008, accessed 2 July 2015).
5. Levine MI and Sackett MF. Results of BCG vaccination in New York City. *Am Rev Tuberculosis* 1946; 53: 517–532.
6. Sackett DL. The tactics of performing therapeutic trials. In: Haynes RB, Sackett DL, Guyatt GH and Tugwell P (eds) *Clinical Epidemiology. How to do clinical practice research*, 4th edn. Philadelphia: Lippincott Williams & Wilkins, 2006, p.86.
7. Drug Project Research Group Coronary. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980; 303: 1038–1041.
8. Furberg CD. How should one analyse and interpret clinical trials in which patients don't take the treatments assigned to them? *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2008, accessed 2 July 2015).
9. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. *J Clin Epidemiol* 2014; 67: 622–628.
10. Cobb LA, Thomas GI, Dillard DH, Merendina KA and Bruce RA. An evaluation of internal-mammary-artery ligation by a double-blind technic. *N Engl J Med* 1959; 260: 1115–1118.
11. Ratcliff JD. New surgery for ailing hearts. *Readers' Digest*, July 1957, p.109–112.
12. Bunker JP, Barnes BA and Mosteller F. *Costs, Risks and Benefits of Surgery*. Oxford: Oxford University Press, 1977.
13. McPherson K and Bunker JP. Costs, risks and benefits of surgery: a milestone in the development of health services research. *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2006, accessed 2 July 2015).
14. Kaptchuk TJ. A brief history of the evolution of methods to control of observer biases in tests of treatments. *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2011, accessed 2 July 2015).
15. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yetisir E and Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology* 1994; 44: 16–20. ([Reprinted as a Classical Article in Neurology 2001; 57(12 Suppl 5): S31–S35]).
16. Colombo F, Shapiro S, Slone D and Tognoni G. *Epidemiological Evaluation of Drugs*. Amsterdam: Elsevier/North-Holland Biomedical Press, 1977.
17. Jick H and Vessey M. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978; 107: 1–7.
18. Ibrahim MA and Spitzer WO eds. *The Case-Control Study: Consensus and Controversy*. New York: Pergamon Press, 1979.
19. Vessey MP. Learning how to control biases in studies to identify adverse effects of drugs. *JLL Bulletin: Commentaries on the history of treatment evaluation*, www.jameslindlibrary.org (2006, accessed 2 July 2015).
20. Sackett DL. Bias in analytic research. *J Chron Dis* 1979; 32: 51–63.

Appendix

The Bermuda conference on 'The Case-Control Study'

It started with a meeting I had with Dave Sackett, Alvan Feinstein and Walter Spitzer (while attending a conference somewhere) in 1978. Dave and I had already been good friends for about 15 years, and a friendship with Alvan and Walter grew out of this meeting. We worked as a 'planning committee' for developing the Bermuda conference, for which I was to serve as chair.

At that time, the atmosphere of epidemiologic research was charged with strong sentiments for and against case-control studies. The 'against group' was small and led by Alvan Feinstein, who gained notoriety as an ardent critic of case-control studies. The focus was on the case-control studies done by Sidney Shapiro, who accessed computerised data to link drugs to health effects. Boehringer Ingelheim was interested, it seemed at that time, in discrediting case-control studies. The company found an 'ally' in Alvan and like-minded people, and consequently gave Walter a grant (no strings attached) to defray the expenses of the conference.

As chair of the conference, I invited about 30 people and selected Bermuda in May as an attractive venue and time for the conference in order to ensure high participation. I did not really know what I was getting into until the opening day, when Alvan and others who disagreed with him, especially Sid Shapiro, began to exchange sharp jibes. I quickly employed whatever expertise I had in diplomacy and in bringing meetings to a successful conclusion in practice and managed to keep everyone civil. (At dinner that evening, Alvan told me that I should negotiate a peace accord between the Jews and the Arabs in the Middle East.) The papers and a discussion (summary) of all the presentations were published both in a special issue of the *Journal of Chronic Diseases* (1979; 32[1 and 2]) [Alvan Feinstein was the editor], and as a book (Ibrahim MA, Spitzer WO, eds (1979). *The*

Case-Control Study: Consensus and Controversy. New York: Pergamon Press).

It was logical to have a presentation on biases that would nicely serve the purpose of the conference. Dave had thought a lot about this issue and had eloquently presented his ideas in various settings. So, it was only natural to ask him to put together a comprehensive presentation on the subject.

The presentation was very well received at the conference and was talked about widely and often since. It was a hit especially among students of epidemiology. I was chair of the University of North Carolina Department of Epidemiology at that time, and I remember that Dave's paper was instrumental in enriching discussions on epidemiologic methods.

Also at that time, Olli Miettinen and Ken Rothman were advancing their own brand of 'new epidemiologic methods'. All of these developments seemed to encourage departments of epidemiology across the country to recruit faculty members whose primary charge was the teaching of 'advanced' and 'new' epidemiologic methods.

Michel Ibrahim, Editor-in Chief, Epidemiologic Reviews, 7 November 2014.

Email: mibrahi1@jhu.edu